$if class loaded beamer$

Unsupervised Learning

Nipun Batra

IIT Gandhinagar

July 29, 2025

Places where you will see unsupervised learning

- It can be used to segment the market based on customer preferences.
- A data science team reduces the number of dimensions in a large dataset to simplify modeling and reduce file size.

REQUIREMENTS: A predefined notion of similarity/dissimilarity.

REQUIREMENTS: A predefined notion of similarity/dissimilarity.
Examples:
Market Segmentation: Customers with similar preferences in the same groups. This would aid in targeted marketing.

$$WCV\left(C_i\right) = \frac{1}{\left|C_i\right|}_{PROTECTED_0}$$

$$\mathrm{WCV}(C_i) = \frac{1}{|C_i|} \sum_{a \in C_i} \sum_{b \in C_i} ||x_a - x_b||_2^2$$

where $|C_i|$ is the number of points in C_i

Then,

$$WCV\left(C_i\right) = \frac{1}{|C_i|} \sum_{a \in C_i} \sum_{b \in C_i} ||x_a - x_b||_2^2$$
$$= 2 \sum_{a \in C_i} ||x_a - x_i||_2^2$$

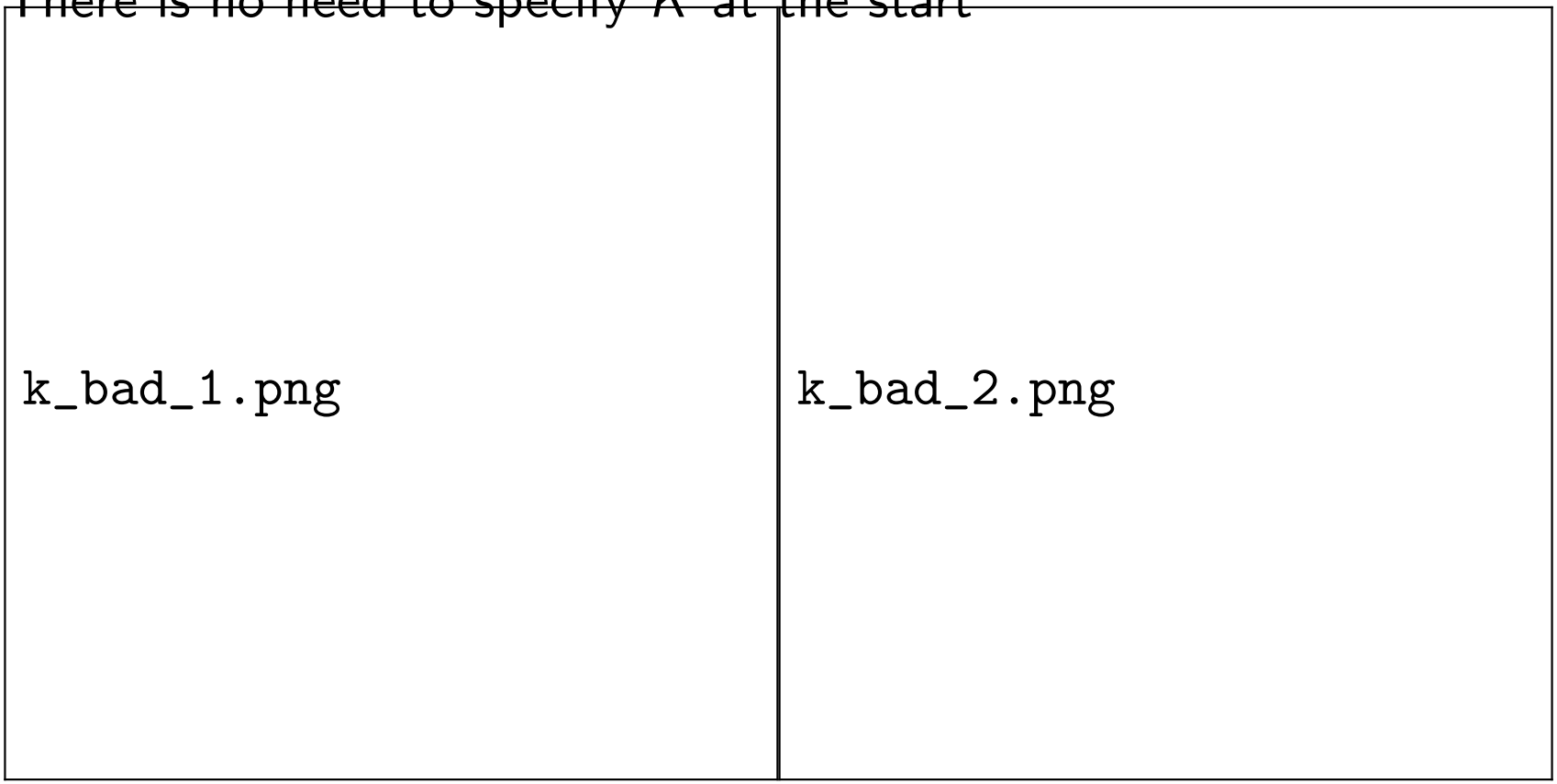
Then,

$$WCV\left(C_i\right) = \frac{1}{|C_i|} \sum_{a \in C_i} \sum_{b \in C_i} ||x_a - x_b||_2^2$$
$$= 2 \sum_{a \in C_i} ||x_a - x_i||_2^2$$

This shows that K-Means gives the **local minima**.

There is no need to specify K at the start

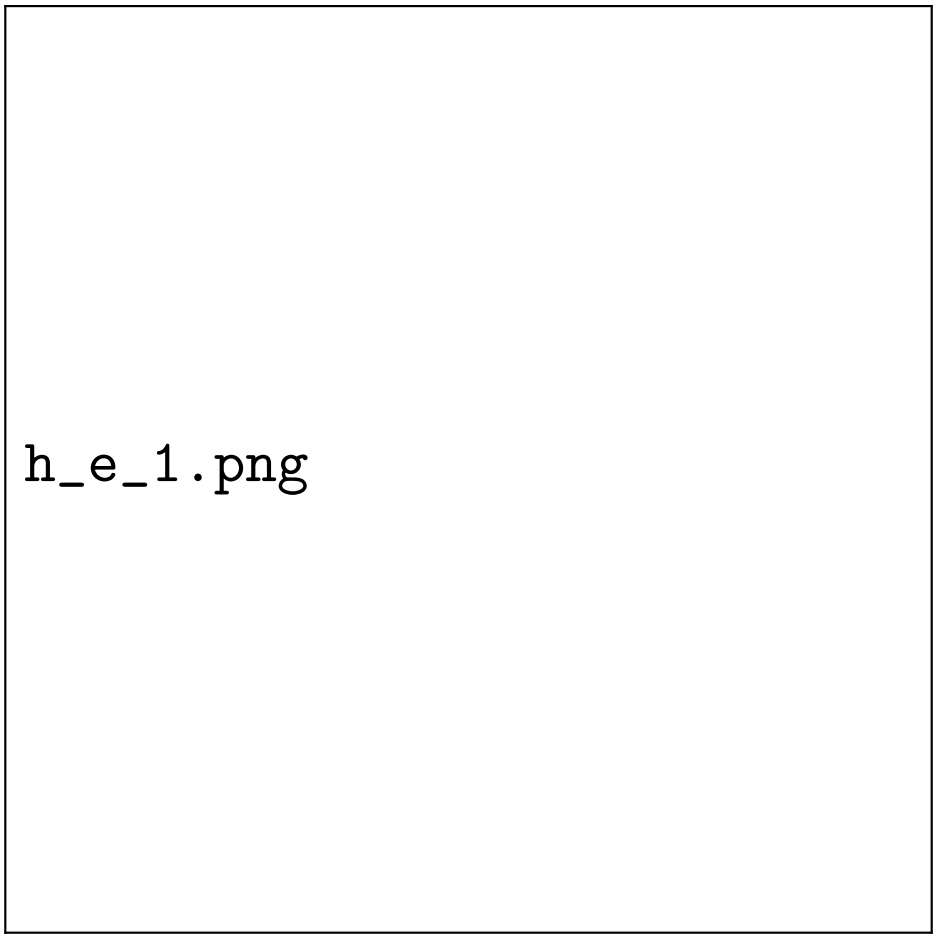
~~There is no need to specify K at the start~~



Examples where K-Means fails

Algorithm for Hierarchal Clustering

1. Start with all points in a single cluster

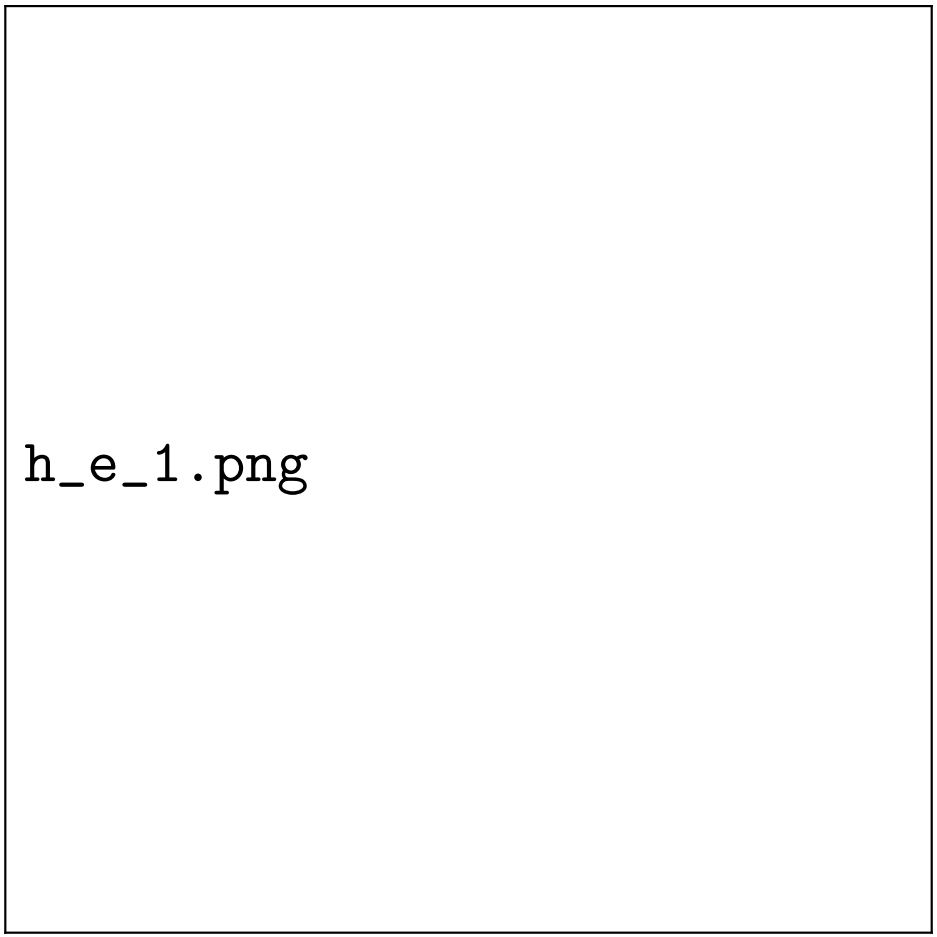


Example Dataset

Algorithm for Hierarchal Clustering

1. Start with all points in a single cluster

   2.1 Identify the 2 closest points

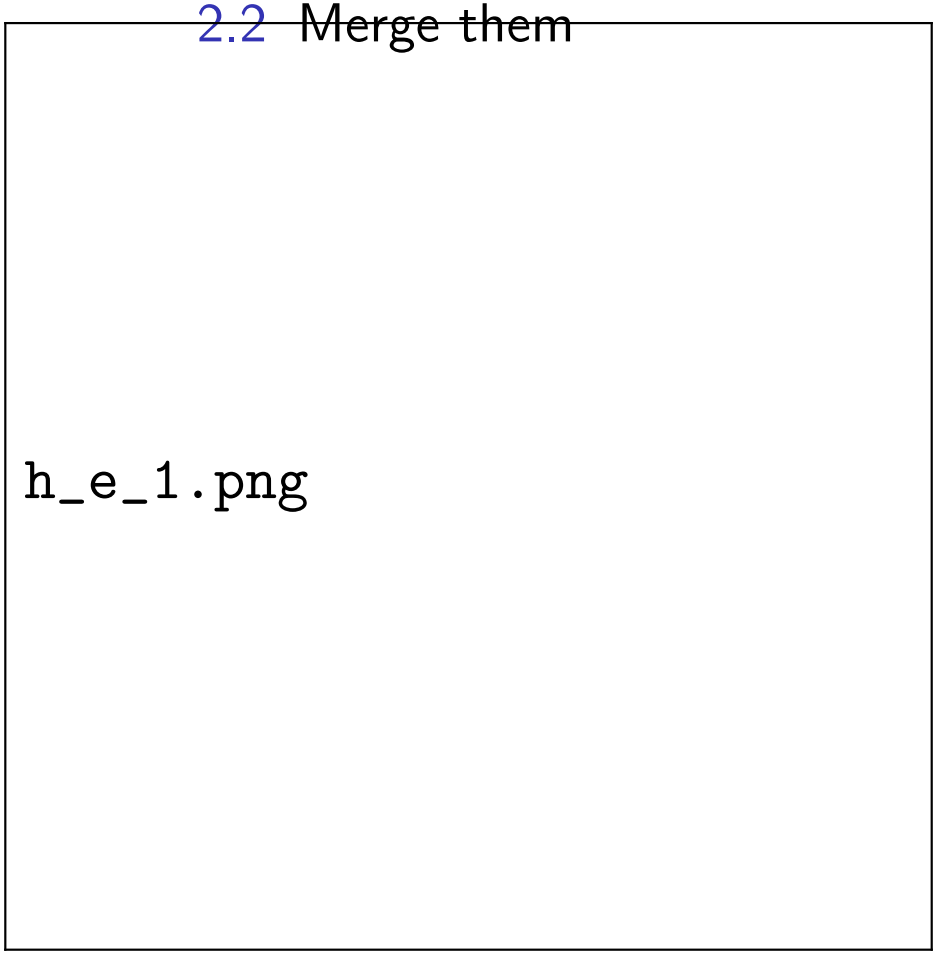


Example Dataset

Algorithm for Hierarchal Clustering

1. Start with all points in a single cluster

   2.1 Identify the 2 closest points
   2.2 Merge them



Example Dataset

Algorithm for Hierarchal Clustering

1. Start with all points in a single cluster
2. Repeat until all points are in a single cluster
   2.1 Identify the 2 closest points
   2.2 Merge them

h_e_1.png

Example Dataset

h_e_2.png

Final Clustering

Joining Clusters/Linkages

Complete
Max inter-cluster similarity

Single
Min inter-cluster similarity

Centroid
Dissimilarity between cluster centroids

More Code

Google Colab Link